Chapter II
Overview of sample design issues for household surveys in developing and transition countries

Ibrahim S. Yansaneh*
International Civil Service Commission
United Nations, New York

Abstract

The present chapter discusses the key issues involved in the design of national samples, primarily for household surveys, in developing and transition countries. It covers such topics as sampling frames, sample size, stratified multistage sampling, domain estimation, and survey analysis. In addition, this chapter provides an introduction to all phases of the survey process which are treated in detail throughout the publication, while highlighting the connection of each of these phases with the sample design process.

Key terms: Complex sample design, sampling frame, target population, stratification, clustering, primary sampling unit.

*Former Chief, Methodology and Analysis Unit, DESA/UNSD.

A. Introduction

1. Sample designs for surveys in developing and transition countries

1. The present chapter presents an overview of issues related to the design of national samples for household surveys in developing and transition countries. The focus, like that of the entire publication, is on household surveys. Business and agricultural surveys are not covered explicitly, but much of the material is also relevant for them.

2. Sample designs for household surveys in developing and transition countries have many common features. Most of the surveys are based on multistage stratified area probability sample designs. These designs are used primarily for frame development and for clustering interviews in order to reduce cost. Sample selection is usually carried out within strata (see sect. B). The units selected at the first stage, referred to in the survey sampling literature as primary sampling units (PSUs), are frequently constructed from enumeration areas identified and used in a preceding national population and housing census. These could be wards in urban areas or villages in rural areas. In some countries, candidates for PSUs include census supervisor areas or administrative districts or subdivisions thereof. The units selected within each selected PSU are referred to as second-stage units, units selected at the third stage are referred to as the third-stage units, and so on. For households in developing and transition countries, second-stage units are typically dwelling units or households, and units selected at the third stage are usually persons. In general, the units selected at the last stage in a multistage design are referred to as the ultimate sampling units.

3. Despite the many similarities discussed above, sample designs for surveys in developing and transition countries are not identical across countries, and may vary with respect to, for example, the target populations, content and objectives, the number of design strata, sampling rates within strata, sample sizes within PSUs, and the number of PSUs selected within strata. In addition, the underlying populations may vary with respect to their prevalence rates for specified population characteristics, the degree of heterogeneity within and across strata, and the distribution of specific subpopulations within and across strata.

2. Overview

4. This chapter is organized as follows. Section A provides a general introduction. Section B considers stratified multistage sample designs. First, sampling with probability proportional to size is described. The concept of design effect is then introduced in the context of cluster sampling. A discussion then follows of the optimum choices for the number of PSUs and the number of second-stage units (dwelling units, households, persons, etc.) within PSUs. Factors taken into consideration in this discussion include the pre-specified precision requirements for survey estimates and practical considerations deriving from the fieldwork organization. Section C discusses sampling frames and associated problems. Some possible solutions to these problems are proposed. Section D addresses the issue of domain estimation and the various allocation schemes that may be considered to satisfy the competing demands arising from the desire to produce estimates at the national and subnational levels. Section E discusses the

determination of the sample size required to satisfy pre-specified precision levels in terms of both the standard error and the coefficient of variation of the estimates. Section F discusses the analysis of survey data and, in particular, emphasizes the fact that appropriate analysis of survey data must take into consideration the features of the sample design that generated the data. Section G provides a summary of some important issues in the design of household surveys in developing and transition countries. A flowchart depicting the important steps involved in a typical survey process, and the interrelationships among the steps of the process, is provided in the annex.

B. Stratified multistage sampling

5. Most surveys in developing and transition countries are based on stratified multistage cluster designs. There are two reasons for this. First, the absence or poor quality of listings of households or addresses makes it necessary to first select a sample of geographical units, and then to construct lists of households or addresses only within those selected units. The samples of households can then be selected from those lists. Second, the use of multistage designs controls the cost of data collection. In the present section, we discuss statistical and operational aspects of the various stages of a typical multistage design.

1. Explicit stratification

6. Stratification is commonly applied at each stage of sampling. However, its benefits are particularly strong in sampling PSUs. It is therefore important to stratify the PSUs efficiently before selecting them.

7. Stratification partitions the units in the population into mutually exclusive and collectively exhaustive subgroups or strata. Separate samples are then selected from each stratum. A primary purpose of stratification is to improve the precision of the survey estimates. In this case, the formation of the strata should be such that units in the same stratum are as homogeneous as possible and units in different strata are as heterogeneous as possible with respect to the characteristics of interest to the survey. Other benefits of stratification include (i) administrative convenience and flexibility and (ii) guaranteed representation of important domains and special subpopulations.

8. Previous sample design and data analysis experience in many countries has pointed to sharp differences in the distribution of population characteristics across administrative regions and across urban and rural areas of each country (see chaps. XXII, XXIII and XXV of this publication for specific examples). This is one of the reasons why, for surveys in these countries, explicit strata are generally based on administrative regions and urban and rural areas within administrative regions. Some administrative regions, such as capital cities, may not have a rural component, while others may not have an urban component. It is advisable to review the frequency distribution of households and persons across these domains before finalizing the choice of explicit sampling strata.

9. In some cases, estimates are desired not only at the national level, but also separately for each administrative region or subregion such as a province, a department or a district. Stratification may be used to control the distribution of the sample based on these domains of interest. For instance, in the Demographic and Health Surveys (DHS) discussed in chapter XXII, initial strata are based on administrative regions for which estimates are desired. Within region, further stratification is effected by urban versus rural components or other types of administrative subdivision. Disproportionate sampling rates are imposed across domains to ensure adequate precision for domain estimates. In general, demand for reliable data for many domains requires large overall sample sizes. The issue of domain estimation in discussed in section D.

2. Implicit stratification

10. Within each explicit stratum, a technique known as implicit stratification is often used in selecting PSUs. Prior to sample selection, PSUs in an explicit strata are sorted with respect to one or more variables that are deemed to have a high correlation with the variable of interest, and that are available for every PSU in the stratum. A systematic sample of PSUs is then selected. Implicit stratification guarantees that the sample of PSUs will be spread across the categories of the stratification variables.

11. For many household surveys in developing and transition countries, implicit stratification is based on geographical ordering of units within explicit strata. Implicit stratification variables sometimes used for PSU selection include residential area (low- income, moderate-income, high-income), expenditure category (usually in quintiles), ethnic group and area of residence in urban areas; and area under cultivation, amount of poultry or cattle owned, proportion of non-agricultural workers, etc., in rural areas. For socio-economic surveys, implicit stratification variables include the proportion of households classified as poor, the proportion of adults with secondary or higher education, and distance from the centre of a large city. Variables used for implicit stratification are usually obtained from census data.

3. Sample selection of PSUs

Characteristics of good PSUs

12. For household surveys in developing and transition countries, PSUs are often small geographical area units within the strata. If census information is available, PSUs may be the enumeration areas identified and used in the census. Similar areas or local population listings are also sometimes utilized. In rural areas, villages may become the PSUs. In urban areas, PSUs may be based on wards or blocks.

13. Since the PSUs affect the quality of all subsequent phases of the survey process, it is important to ensure that the units designated as PSUs are of good quality and that they are selected for the survey in a reasonably efficient manner. For PSUs to be considered of good quality, they must, in general:

(a) Have clearly identifiable boundaries that are stable over time;

(b) Cover the target population completely;

(c) Have a measure of size for sampling purposes;

(d) Have data for stratification purposes;

(e) Be large in number.

14. Before sample selection, the quality of the sampling frame needs to be evaluated. For a frame of enumeration areas, a first step is to review census counts by domains of interest. In general, considerable attention should be given to the nature of the PSUs and the distribution of households and individuals across the PSUs for the entire population and for the domains of interest. A careful examination of these distributions will inform decisions about the choice of PSU and will identify units that need adjustment in order to conform to the specifications of a good PSU. In general, a wide variability in the number of households and persons across PSUs and across time would have an adverse effect on the fieldwork organization. If the PSUs are selected with equal probability, it would also have an adverse effect on the precision of survey estimates.

15. Often, natural choices for PSUs are not usable because they are deficient in the sense that they lack one or more of the above features. Such PSUs need to be modified or adjusted before they are used. For instance, if the boundaries of enumeration areas are thought to be not well defined, then larger and more clearly defined units such as administrative districts, villages, or communes may be used as PSUs. Furthermore, PSUs considered to be extremely large are sometimes split or alternatively treated as strata, often known as certainty selections or "self-representing" PSUs (see Kalton, 1983). Small PSUs are usually combined with neighbouring ones in order to satisfy the requirement of a pre-specified minimum number of households per PSU. The adjustment of under and oversized PSUs is best carried out prior to sample selection.

16. To ensure an equitable distribution of sampled households within PSUs, very large PSUs are sometimes partitioned into a number of reasonably sized sub-units, one of which is randomly selected for further field operations, such as household listing. This is called *chunking* or *segmentation*. Note that the selection and segmentation of oversized PSUs introduce an extra stage of sampling, which must be accounted for in the weighting process.

17. Very small PSUs can also be combined with neighbouring PSUs on the PSU frame in order to satisfy a pre-specified minimum measure of size for PSUs. However, the labour involved in combining small PSUs is considerably reduced by carrying out the grouping either during or after the selection of PSUs. However, this is a tedious process requiring adherence to strict rules and a lot of record keeping. A procedure for combining PSUs during or after sample selection is described in Kish (1965). One disadvantage of this procedure is that it does not guarantee that the PSUs selected for grouping are contiguous. Therefore, this procedure is not recommended in situations where the number of undersized PSUs is large.

Problems with inaccurate measures of size and possible solutions

18. One of the most common problems with frames of enumeration areas that are used as PSUs - as is typically done in developing and transition countries - is that the measures of size may be very inaccurate. The measures of size are generally counts of numbers of persons or households in the PSUs based on the last population census. They may be significantly out of date, and they may be markedly different from the current sizes because of such factors as growth in urban areas and shrinkage in other areas as a result of migration, wars, and natural disasters. Inaccurate measures of size lead to lack of control over the distribution of second-stage units and the sub-sample sizes, and this can cause serious problems in subsequent field operations. One solution to the problem of inaccurate measures of size is to conduct a thorough listing operation to create a frame of households in selected PSUs before selecting households. Another solution is to select PSUs with probability proportional to estimated size. Both of these procedures are elaborated in sections 4 and 5 below. Other common problems associated with using enumeration areas as PSUs include the lack of good-quality maps and incomplete coverage of the target population, one of several sampling frame-related problems discussed in section C.

4. Sampling of PSUs with probability proportional to size

19. Prior to sample selection, PSUs are stratified explicitly and implicitly using some of the variables listed in sections B.1 and B.2. For most household surveys in developing and transition countries, PSUs are selected with probability proportional to a measure of size. Before sample selection, each PSU is assigned a measure of size, usually based on the number of households or persons recorded for it during a recent census or as the result of a recent updating exercise. Then, a separate sample of PSUs is selected within each explicit stratum with probability proportional to the assigned measure of size.

20. Probability proportional to size (PPS) sampling is a technique that employs auxiliary data to yield dramatic increases in the precision of survey estimates, particularly if the measures of size are accurate and the variables of interest are correlated with the size of the unit. It is the methodology of choice for sampling PSUs for most household surveys. PPS sampling yields unequal probabilities of selection for PSUs. Essentially, the measure of size of the PSU determines its probability of selection. However, when combined with an appropriate subsampling fraction for selecting households within selected PSUs, it can lead to an overall self-weighting sample of households in which all households have the same probability of selection regardless of the PSUs in which they are located. Its principal attraction is that it can lead to approximately equal sample sizes per PSU.

21. For household surveys, a good example of a PPS size variable for the selection of PSUs is the number of households. Admittedly, the number of households in a PSU changes over time and may be out of date at the time of sample selection. However, there are several ways of dealing with this problem, as discussed in paragraph 18. For farm surveys, a PPS size measure that is frequently used is the size of the farm. This choice is in part because typical parameters of interest in farm surveys, such as income, crop production, livestock holdings and expenses are correlated with farm size. For business surveys, typical PPS measures of size include the number of employees, number of establishments and annual volume of sales. Like the number

of households, these PPS measures of size are likely to change over time, and this fact must be taken into consideration in the sample design process.

22. Consider a sample of households, obtained from a two-stage design, with *a* PSUs selected at the first stage and a sample of households at the second stage. Let the measure of size (for example, the number of households at the time of the last census) of the i^{th} PSU be M_i. If the PSUs are selected with PPS, then the probability P_i of selecting the i^{th} PSU is given by

$$P_i = a \times \frac{M_i}{\sum_i M_i}$$

23. Now, let $P_{j|i}$ denote the conditional probability of selecting the j^{th} household in the i^{th} PSU, given that the i^{th} PSU was selected at the first stage. Then, the selection equation for the unconditional probability P_{ij} of selecting the j^{th} household in the i^{th} PSU under this design is

$$P_{ij} = P_i \times P_{j|i}$$

24. If an equal-probability sample of households is desired with an overall sampling fraction of $f = P_{ij}$, then households must be selected at the appropriate rate, inversely proportional to the probability of selection of the PSUs in which they are located, that is to say,

$$P_{j|i} = \frac{f}{P_i}$$

25. If the measures of size of the PSUs are the true sizes, and there is no change in the measure of size between sample selection and data collection, and if *b* households are selected in each sampled PSU, then we obtain a self-weighting sample of households with a probability of selection given by

$$P_{ij} = a \times \frac{M_i}{\sum_i M_i} \times \frac{b}{M_i} = \frac{a \times b}{\sum_i M_i} = f$$

where *f* is a constant.

26. The problem with this procedure is that the true measures of size are rarely known in practice. However, it is often possible to obtain good estimates, such as population and household counts from a recent census, or some other reliable source. This allows us to apply the procedure known as probability-proportional-to-estimated-size (PPES) sampling. There are two choices for PPES sampling in a two-stage design with households selected at the second stage: either (a) select households at a fixed rate in each sampled PSU; or (b) select a fixed number of households per sampled PSU.

27. PPES sampling of households at a fixed rate is implemented as follows. Let the true values of the measure of size be denoted by N_i, and assume that the values M_i are good estimates of N_i. We then apply the sampling rate b/M_i to the i^{th} PSU to obtain a sample size of

$$b_i = \frac{b}{M_i} \times N_i$$

28. Note that subsampling within PSUs at a fixed rate (inversely proportional to the measures of size of the PSUs) involves the determination of a rate for each sampled PSU so that, together with the PSU selection probability, we obtain an equal-probability sample of households, regardless of the actual size of the PSUs. However, this procedure does not provide control over the subsample sizes, and hence the overall sample size. More households will be sampled from PSUs with larger-than-expected numbers of households, and fewer households will be sampled from PSUs with smaller-than-expected numbers of households. This has implications for the fieldwork organization. In addition, if the measures of size are so out of date that the variation in the realized samples is extreme, there may be a need for a change in the sampling rate so as to obtain sample sizes that are a bit more homogeneous across PSUs, which would entail some degree of departure from a self-weighting design.

29. The second procedure, selecting a fixed number of households per PSU, avoids the disadvantage of variable sample sizes per PSU but does not produce a self-weighting sample. However, if the measures of size are updated immediately prior to sample selection of PSUs, they may provide good enough approximations that will lead to an approximately self-weighting sample of households.

30. In summary, even though subsampling within PSUs at a fixed rate is designed to produce self-weighting samples, there are circumstances under which this method leads to departures from a self-weighting sample of households. On the other hand, even though selecting a fixed number of households within PSUs often does not produce self-weighting samples, there are circumstances under which this method leads to approximately self-weighting samples of households. Whenever there are departures from a self-weighting design, weights must be used to compensate for the resulting differential selection probabilities in different PSUs.

5. Sample selection of households

31. Once the sample selection of PSUs is completed, a procedure is carried out whose aim is to list all households or all housing units or dwellings in each selected PSU. Sometimes the listings are of dwelling units and then all households in selected dwelling units are included if a dwelling unit is sampled. The objective of this listing step is to create an up-to-date sampling frame from which households can be selected. The importance of carrying out this step effectively cannot be overemphasized. The quality of the listing operation is one of the most important factors that affect the coverage of the target population.

32. Prior to sample selection in each sampled PSU, the listed households may be sorted with respect to geography and other variables deemed strongly correlated with the survey variables of

interest (see sect. B.2). Then, households are sampled from the ordered list by an equal-probability systematic sampling procedure. As indicated in section B.4, households may be selected within sampled PSUs at sampling rates that generate equal overall probabilities of selection for all households or at rates that generate a fixed number of sampled households in each PSU. The merits and demerits of these approaches are discussed in section B.4.

33. Frequently, the ultimate sampling units are households and information is collected on the selected households and all members of those households. For special modules covering incomes and expenditures, for which households are the units of analysis, a knowledgeable respondent is often selected to be the household informant. For subjects considered sensitive for persons within households (for example, domestic abuse), a random sample of persons (frequently of one person) is selected within each sampled household.

6. Number of households to be selected per PSU

34. Primary sampling units consist of sets of households that are geographically clustered. As a result, households in the same cluster generally tend to be more alike in terms of the survey characteristics (for example, income, education, occupation, etc.) than households in general. Clustering reduces the cost of data collection considerably, but correlations among units in the same cluster inflate the variance (lower the precision) of survey estimates, compared with a design in which households are not clustered. Thus the challenge for the survey designer is to achieve the right balance between the cost savings and the corresponding loss in precision associated with clustering.

35. The inflation in variance of survey estimates attributable to clustering contributes to the so-called design effect. The design effect represents the factor by which the variance of an estimate based on a simple random sample of the same size must be multiplied to take account of the complexities of the actual sample design due to stratification, clustering and weighting. It is defined as the ratio of the variance of an estimate based on the complex design relative to that based on a simple random sample of the same size. See chaps. VI and VII of this publication, and the references cited therein, for details on design effects and their use in sample design. An expression for the design effect (due to clustering) for an estimate [for example, an estimated mean ($\bar{y}$)] is given approximately by:

$$D^2(\bar{y}) = 1 + (b-1)\rho$$

where $D^2(\bar{y})$ denotes the design effect for the estimated mean ($\bar{y}$), ρ is the intra-class correlation, and b is the average number of households to be selected from each cluster, that is to say, the average cluster sample size. The intra-class correlation is a measure of the degree of homogeneity (with respect to the variable of interest) of the units within a cluster. Since units in the same cluster tend to be similar to one another, the intra-class correlation is almost always positive. For human populations, a positive intra-class correlation may be due to the fact that households in the same cluster belong to the same income class; may share the same attitudes towards the issues of the day; and are often exposed to the same environmental conditions (climate, infectious diseases, natural disaster, etc.).

36. Failure to take account of the design effect in the estimates of standard errors can lead to invalid interpretation of the survey results. It should be noted that the magnitude of $D^2(\bar{y})$ is directly related to the value of *b*, the cluster sample size, and the intra-class correlation (ρ). For a fixed value of ρ, the design effect increases linearly with *b*. Thus, to achieve low design effects, it is desirable to use as small a cluster sample size as possible. Table II.1 illustrates how the average cluster size and the intra-class correlation affect the design effect. For example, with an average cluster sample size *b* of 20 dwelling units per PSU and ρ equal to 0.05, the design effect is 1.95. In other words, this cluster sample design yields estimates with the same variance as those from an unclustered (simple random) sample of about half the total number of households. With larger values of ρ, the loss in precision is even greater, as can be seen on the right-hand side of table II.1.

Table II.1. Design effects for selected combinations of cluster sample size and intra-class correlation

Cluster	Intra-class correlation (ρ)								
Sample size (*b*)	0.005	0.01	0.02	0.03	0.04	0.05	0.10	0.20	0.30
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	1.05	1.09	1.18	1.27	1.36	1.45	1.90	2.80	3.70
15	1.07	1.14	1.28	1.42	1.56	1.70	2.40	3.80	5.20
20	1.10	1.19	1.38	1.57	1.76	1.95	2.90	4.80	6.70
30	1.15	1.29	1.58	1.87	2.16	2.45	3.90	6.80	9.70
50	1.25	1.49	1.98	2.47	2.96	3.45	5.90	10.80	15.70

37. In general, the optimum number of households to be selected in each PSU will depend on the data-collection cost structure and the degree of homogeneity or clustering with respect to the survey variables within the PSU. Assume a two-stage design with PSUs selected at the first stage and households selected at the second stage. Also, assume a linear cost model for the overall cost related to the sampling of PSUs and households given by

$$C = aC_1 + abC_2$$

where C_1 and C_2 are, respectively, the cost of an additional PSU and the cost of an additional household; and *a* and *b* denote, respectively, the number of selected PSUs and the number of households selected per PSU (Cochran, 1977, p. 280). Under this cost model, the optimum choice for *b* that minimizes the variance of the sample mean (see Kish, 1965, sect. 8.3.b) is approximately given by

$$b_{opt} = \sqrt{\frac{C_1}{C_2}\frac{(1-\rho)}{\rho}}\,.$$

38. Table II.2 gives the optimal subsample size (*b*) for various cost ratios C_1/C_2 and intra-class correlation. Note that all other things being equal, the optimal sample size decreases (that

is to say, the sample is more broadly spread across clusters) as the intra-class correlation increases and as the cost of an additional household increases relative to that of a PSU.

39. The cost model used in the derivation of the optimal cluster size is an oversimplified one but is probably adequate for general guidance. Since most surveys are multi-purpose in nature, involving different variables and correspondingly different values of ρ, the choice of b often involves a degree of compromise among several different optima.

Table II.2. Optimal subsample sizes for selected combinations of cost ratio and intra-class correlation

Cost ratio	Intra-class correlation				
(C_1/C_2)	0.01	0.02	0.03	0.05	0.08
4	20	14	11	9	5
9	30	21	17	13	10
16	40	28	23	17	14
25	50	35	28	22	17

40. In the absence of precise cost information, table II.2 can be used to determine the optimal number households to be selected in a cluster for various choices of cost ratio and intra-class correlation. For instance, if it is known a priori that the cost of including a PSU is four times as great as that of including a household, and that the inter-class correlation for a variable of interest is 0.05, then it is advisable to select about nine households in the cluster. Note that the optimum number of households to be selected in a cluster does not depend on the overall budget available for the survey. The total budget determines only the number of PSUs to be selected.

41. In general, the factors that need to be considered in determining the sample allocation across PSUs and households within PSUs include the precision of the survey estimates (through the design effect), the cost of data collection and the fieldwork organization. If travel costs are high, as is the case in rural areas, it is preferable to select a few PSUs and many households in each PSU. On the other hand, if, as in urban areas, travel costs are lower, then it is more efficient to select many PSUs and, then, fewer households within each PSU. On the other hand, in rural areas, it may be more efficient to select more households per PSU. These choices must be made in such a way as to produce an efficient distribution of workload among the interviewers and supervisors.

C. Sampling frames

1. Features of sampling frames for surveys in developing and transition countries

42. For most household surveys, the target population comprises the civilian non-institutionalized population. In order to obtain the desired data from this target population, interviews are often conducted at the household level. In general, only persons considered permanent residents of the household are eligible for inclusion in the surveys. Permanent residents of a household who are away temporarily, such as persons on vacation, or temporarily

in a hospital, and students living away from home during the school year, are generally included if their household is selected. Students living away from home during the school year are not included in the survey if sampled at their school-time residence because data for such students would be obtained from their permanent place of residence. Groups that are generally excluded from household surveys in developing and transition countries include members of the armed forces living in barracks or in private homes; persons in prisons, hospitals, nursing homes or other institutions; homeless people; and nomads. Most of these groups are generally excluded because of the practical difficulties usually encountered in collecting data from them. However, the decision on whether or not to exclude a group needs to be made in the light of the survey objectives.

2. Sampling frame problems and possible solutions

43. As in other types of surveys, the quality of data obtained from household surveys depends to a large extent on the quality of the sampling frame from which the sample for the survey was selected. Unfortunately, problems with sampling frames are an inevitable feature of household surveys. The present section discusses some of these problems and suggests possible solutions.

44. Kish (1965, sect. 2.7) provides a useful classification of four frame problems and possible solutions for them. The four problems are non-coverage, clusters of elements, blanks, and duplicate listings. We discuss these errors in the context of multistage designs for surveys conducted in developing and transition countries.

45. The term "non-coverage" refers to the failure of the sampling frame to cover all of the target population, as a result of which some sampling units have no probability of inclusion in the sample. Non-coverage is a major concern for household surveys conducted in developing and transition countries. Evidence of the impact of non-coverage can be seen from the fact that sample estimates of population counts based on most surveys in developing and transition countries fall well short of population estimates from other sources.

46. There are three levels of non-coverage: the PSU level, the household level and the person level. For developing and transition countries, non-coverage of PSUs is a less serious problem than non-coverage of households and of eligible persons within sampled households. Non-coverage of PSUs occurs, for example, when some regions of a country are excluded from a survey on purpose, because they are inaccessible, owing to war, natural disaster or other causes. Also, remote areas with very few households or persons are sometimes removed from the sampling frames for household surveys because they represent a small proportion of the population and so have very little effect on the population figures. Non-coverage is a more serious problem at the household and person levels. Households or persons may be erroneously excluded from the survey as the result of the complex definitional and conceptual issues regarding household structure and composition. There is potential for inconsistent interpretation of these issues by different interviewers or those responsible for creating lists of households and household members. Therefore, strict operational instructions are needed to guide interviewers on who is to be considered a household member and on what is to be considered a household or a dwelling unit. As a means of addressing this problem, the quality of the listing of households

and eligible persons within households should be made a key area for methodological work and training in developing and transition countries.

47. The problem of blanks arises when some listings on the sampling frame contain no elements of the target population. For a list frame of dwelling units, a blank would correspond to an empty dwelling. This problem also arises in instances where one is sampling particular subgroups of the population, for instance, women who had given birth last year. Some households that were listed and sampled will not contain any women who gave birth last year. If possible, blanks can be removed from the frame before sample selection. However, this is not cost-effective in many practical applications. A more practical solution is to identify and eliminate blanks after sample selection. However, eliminating blanks means that the realized sample will be smaller and of variable size.

48. The problem of duplicate listings arises when units of the target population appear more than once in the sampling frame. This problem can arise, for example, when one is sampling nomads or part-year residents in one location. One way to avoid duplicate listings is to designate a pre-specified unique listing as the actual listing and the other listings as blanks. Only if the unique listing is sampled is the unit included in the sample. For example, nomads who herd their cattle in moving from place to place in search of grazing land and water for their animals may be sampled as they go to the watering holes. Depending on the drinking cycles of the animals (horses reportedly have longer cycles that cattle), some are likely to visit more than one watering hole in the survey data-collection period. To avoid duplicate listings, nomads might be uniquely identified with their first visit to a watering hole after a given date, with later visits being treated as blanks. Otherwise, the weights of the sampled units need to be adjusted to account for the duplicates. See Yansaneh (2003) for examples of how this is done.

49. The problem of clusters of elements arises when a single listing on the sampling frame actually consists of multiple units in the target population. For example, a list of dwellings may contain some dwellings with more than one household. In such instances, the inclusion of all households linked to the sampled dwelling will yield a sample in which the households have the same probability of selection as the dwelling. Note that the practice of randomly selecting one of the units in the cluster automatically leads to unequal probabilities of selection, which would need to be compensated for by weighting.

3. Maintenance and evaluation of sampling frames

50. The construction and maintenance of good sampling frames constitute an expensive and time-consuming exercise. Developing and transition countries have the potential to create such frames from such sources as decennial census data. It is advisable that every national statistics office set as a high priority the creation and maintenance of a master sampling frame of enumeration areas that were defined and used in a preceding census. Such a sampling frame should be established soon after the completion of the census, because the amount of labour involved increases with the distance in time from the census. The frame must have appropriate labels of other, possibly larger, geographical areas that may be used as primary sampling units. It should also include data that may be useful for stratification, such as ethnic and racial composition, median expenditure or expenditure quintiles, etc. If properly maintained, the

master sampling frame can be used to service an integrated system of surveys including repeated surveys. See chapter V for details about the construction and maintenance of master sampling frames.

D. Domain estimation

1. Need for domain estimates

51. In recent years, there has been increasing demand in most countries for reliable data not only at the national level, but also for subnational levels or domains, owing mainly to the fact that most development or intervention programmes are implemented at subnational levels, such as that of the administrative region or the district. Making important decisions concerning programme implementation or resource allocation at the local level requires precise data at that level.

52. For the purposes of this discussion, we will define a domain as any subset of the population for which separate estimates are planned in the survey design. A domain could be a stratum, a combination of strata, an administrative region, or urban, rural or other subdivisions within these regions. For example, estimates from many national surveys are published separately for administrative regions. The regions can then be treated as domains, each with two strata (for example, urban and rural subpopulations) or more. Domains can also be demographic subpopulations defined by such characteristics as age, race and sex. However, a complication arises when the domains cut across stratum boundaries, as in the case, for instance, where a domain consists of households with access to health services.

53. It is important that the number of domains of interest for a particular survey be kept at a moderate level. The sample size required to provide reliable estimates for each of a large number of domains would necessarily be very large. The problems associated with large samples will be discussed in section E.

2. Sample allocation

54. Provision of precise survey estimates for domains of interest requires that samples of adequate sizes be allocated to the domains. However, conflicts arise when equal precision is desired for domains with widely varying population sizes. If estimates are desired at the same level of precision for all domains, then an equal allocation (that is to say, the same sample size per domain) is the most efficient strategy. However, such an allocation can cause a serious loss of efficiency for national estimates. Proportionate allocation, which uses equal sampling fractions in each domain, is frequently the most suitable allocation for national estimates. When domains differ markedly in size and when both national and domain estimates are required, some compromise between equal allocation and equal sampling fractions is required.

55. A compromise between proportional and equal allocation was proposed by Kish (1988), based on an allocation proportional to $n\sqrt{(W_h^2 + H^{-2})}$, where n is the overall sample, size, W_h is the proportion of the population in stratum h and H is the number of strata. For very small strata,

the second term dominates the first, thereby preventing allocations to the small strata that are too small.

56. An alternative approach is to augment the sample sizes of smaller domains to the extent necessary to satisfy the required precision levels. When a domain is small, proportional allocation will yield a sample size for the domain that may be too small to generate sufficiently precise estimates. The remedy is to oversample, or sample at a higher rate, from the small domains.

57. To summarize, survey designers in developing and transition countries are often confronted with the choice between precise estimates at the national level and precise estimates for the domains. This problem becomes more serious when the domains of interest have widely varying sizes. One way to circumvent this dilemma is to define domains that are approximately equal in size, perhaps by combining existing domains. Alternatively, the domains can be kept distinct and a lower precision level may be allowed for the small domains or, perhaps, there will be no estimates published for the domains.

E. Sample size

1. Factors that influence decisions about sample size

58. Both producers and users of survey data often desire large sample sizes because they are deemed necessary to make the sample more "representative", and also to minimize sampling error and hence increase the reliability of the survey estimates. This argument is advanced almost without regard to the possible increase in non-sampling errors that comes from large sample sizes. In the present section, we discuss the factors that must be taken into consideration in determining the appropriate sample size for a survey.

59. The three major issues that drive decisions about the appropriate sample size for a survey are:

- Precision (reliability) of the survey estimates
- Quality of the data collected by the survey
- Cost in time and money of data collection, processing and dissemination

We now discuss each of these factors in turn.

2. Precision of survey estimates

60. The objectives of most surveys in developing and transition countries include the estimation of the level of a characteristic (for instance, the proportion of households classified as poor), at a point in time and of the change in that level over time (for instance, the change in the poverty rate between two points in time). We discuss the precision of survey estimates in the context of estimation of the level of a characteristic at a point in time. For the rest of the

discussion, we will use the percentage of households in poverty, which we will call the poverty rate, as the characteristic of interest.

61. The precision of an estimate is measured by its standard error. The formula for the estimated standard error of an estimated poverty rate *p* in a given domain, denoted by *se(p)*, is given by

$$se(p) = \sqrt{d^2(p) \times (1 - \frac{n}{N}) \times \frac{p(100-p)}{n}}$$

where *n* denotes the overall number of households for the domain of interest, *N* denotes the total number of households in the domain and $d^2(p)$ denotes the estimated design effect associated with the complex design of the survey.[2] The proportion of the population that is in the sample, *n/N*, is called the sampling fraction and the factor $[1-(n/N)]$ (the proportion of the population not included in the sample), is called the finite population correction factor (fpc). The fpc represents the adjustment made to the standard error of the estimate to account for the fact that the sample is selected without replacement from a finite population.

62. We will use data from Viet Nam for illustration. The total number of households, *N*, based on the 1999 population census is 16,661,366. See Glewwe and Yansaneh (2000) for details on the distribution of households based on the 1999 census. Note that, with such a large population size, the finite population correction factor is negligible in all cases. Table II.3 provides standard errors and 95 per cent confidence intervals for various estimates of the poverty rate, assuming a design effect of 2.0. A 95 per cent confidence interval is one with a 95 per cent probability of containing the true value. The table shows that for a given sample size, the standard errors increase as the poverty rate increases, reaching a maximum for p = 50 per cent. The associated 95 per cent confidence intervals also become wider with an increasing poverty rate, being the widest when the poverty rate is 50 per cent. Thus, in general, domains with poverty rates much smaller or larger than 50 per cent will have more precise survey estimates relative to domains with poverty rates near 50 per cent, for a given sample size and design effect.[3] This means that domains with very low or very high rates of poverty will require a smaller sample size to achieve the same standard error as a domain with a poverty rate close to 50 per cent. For example, consider a sample size of 500 households in a domain. If such a domain has an estimated poverty rate of only 5 per cent, the confidence interval is 5 ± 2.7 per cent; if the domain has an estimated poverty rate of 10 per cent, the confidence interval is 10 ± 3.7 per cent; if the domain has an estimated poverty rate of 25 per cent, the confidence interval is 25 ± 5.4 per cent; and if the domain has an estimated poverty rate of 50 per cent, the confidence interval is 50 ± 6.2 per cent.

---

[2] Although *n* should actually be *n-1* in the above formula for *se(p)*, in most practical applications, *n* is large enough for the difference between *n* and *n-1* to be negligible.

[3] For poverty rates of greater than 50 per cent (p > 50 per cent), the standard error is the same as that for a poverty rate of 100 – p, and thus can be inferred from Table III.3. For example, the standard error of an estimated poverty rate of 75 per cent is the same as that of an estimated poverty rate of 25 per cent.

Table II.3. Standard errors and confidence intervals for estimates of poverty rate based on various sample sizes, with the design effect assumed to be 2.0

	Poverty rate (percentage)									
	5		10		25		40		50	
Sample size	Standard error	Confidence interval	Standard error	Confidence interval	Standard error	Confidence interval	Standard error	Confidence interval	Standard error	Confidence Interval
250	1.95	(1.2 , 8.8)	2.68	(4.7 , 15.3)	3.87	(17.4 , 32.6)	4.38	(31.4 , 48.6)	4.47	(41.2 , 58.8)
500	1.38	(2.3 , 7.7)	1.90	(6.3 , 13.7)	2.74	(19.6 , 30.4)	3.10	(33.9 , 46.1)	3.16	(43.8 , 56.2)
750	1.13	(2.8 , 7.2)	1.55	(7.0 , 13.0)	2.24	(20.6 , 29.4)	2.53	(35.0 , 45.0)	2.58	(44.9 , 55.1)
1000	0.97	(3.1 , 6.9)	1.34	(7.4 , 12.6)	1.94	(21.2 , 28.8)	2.19	(35.7 , 44.3)	2.24	(45.6 , 54.4)
1500	0.80	(3.4 , 6.6)	1.10	(7.9 , 12.1)	1.58	(21.9 , 28.1)	1.79	(36.5 , 43.5)	1.83	(46.4 , 53.6)
2000	0.44	(4.1 , 5.9)	0.95	(8.1 , 11.9)	1.37	(22.3 , 27.7)	1.55	(37.0 , 43.0)	1.58	(46.9 , 53.1)

63. Of course, increasing the sample size to more than 500 households reduces the width of the confidence interval (in other words, the sample estimate becomes more precise). However, the reduction in width is proportional not to the increase in sample size, but to the square root of that increase, in this case $\sqrt{n/500}$, where *n* is the new sample size. For example, in a domain with a poverty rate of 25 per cent, doubling the sample size from 500 to 1,000 households would reduce the width of the confidence interval by a factor of $\sqrt{2}$, that is to say, from ± 5.4 per cent to ± 3.8 per cent. Such reductions should be carefully weighed against the increased complexities in the management of survey operations, survey costs and non-sampling errors.

64. The precision of survey estimates is often expressed in terms of the coefficient of variation of the estimate of interest. As before, we restrict attention to the estimation of the percentage of households classified as poor in a country. The estimated coefficient of variation of an estimate of the poverty rate, denoted by *cv(p)*, is given by

$$cv(p) = \frac{se(p)}{p} = \sqrt{d^2(p) \times (1 - \frac{n}{N}) \times \frac{(100 - p)}{np}}$$

65. Table II.4 presents the estimated coefficients of variation for an estimated poverty rate for various sample sizes, assuming a design effect of 2.0, where *cv* is expressed as a percentage. The table shows that for a given sample size, the estimated coefficient of variation of the estimated poverty rate decreases steadily as the true percentage increases. Also, for a given poverty rate, the coefficient of variation decreases as the sample size decreases. For a sample size of 500, the coefficient of variation is about 28 per cent when p = 5 per cent, 19 per cent when p = 10 per cent, 11 per cent when p = 25 per cent, 8 per cent when p = 40 per cent, 6 per

cent when p = 50 per cent, 5 per cent when p = 60 per cent, 4 per cent when p = 75 per cent, 2 per cent when p = 90 per cent, and 1 per cent when p = 95 per cent. As the sample size increases, the estimated coefficient of variation decreases correspondingly. Note that unlike the standard errors shown in table II.3, the coefficient of variation shown in table II.4 is not a symmetric function of the poverty rate.

Table II.4. Coefficient of variation for estimates of poverty rate based on various sample sizes, with the design effect assumed to be 2.0

	Poverty rate (percentage)								
Sample size	5	10	25	40	50	60	75	90	95
250	39	27	15	11	9	7	5	3	2
500	28	19	11	8	6	5	4	2	1
750	23	15	9	6	5	4	3	2	1
1000	19	13	8	5	4	4	3	1	1
1500	16	11	6	4	4	3	2	1	1
2000	14	9	5	4	3	3	2	1	1

3. Data quality

66. An important consideration in the determination of the sample size for a survey is the quality of the data that will be collected. It is important to maintain data of the highest possible quality so that one can have confidence in the estimates generated from them. Checking the quality of the data at every stage of the implementation of the survey is essential. As a result, it is important to keep the sample size to a reasonable limit so that adequate checking and editing can be done in a fashion that is efficient in terms of both time and money.

67. A factor related to sample size that affects data quality is the number of staff working on the study. For instance, smaller sample sizes require fewer interviewers, so that these interviewers can be more selectively chosen. In particular, with a smaller sample size, it is more likely that all interviewers will be recruited from the ranks of well-trained and experienced staff. Moreover, interviewers will be better trained because with a small number of interviewers, the training can be better focused and proportionately more survey resources can be devoted to it. Fewer training materials will be needed and interviewers will receive more individual attention during training and in the field. All of this will result in fewer problems in data collection and in subsequent editing of the data collected. Consequently, the data available for analysis will be of a higher quality, permitting policy makers to have greater confidence in the decisions being made on the basis of these data.

68. In addition to concerns about the quality of the data collected, larger sample sizes make it more difficult and expensive to minimize survey non-response (see chap. VIII). It is important to keep survey non-response as low as possible, in order to reduce the possibility of large biases in the survey estimates (see sect. F.1). Such biases could result if we fail to secure responses from a sizeable portion of the population that may be considerably different from those included in the survey. For example, persons who live in urban areas and have relatively high incomes

are often less likely to participate in household surveys. Failure to include a large segment of this portion of the population can lead to the underestimation of such population characteristics as the national average household income, educational attainment and literacy. With a smaller sample, it will be much easier and more cost-effective to revisit households that initially chose not to participate, in an attempt to persuade them to do so. Since persuading initial non-participants to become participants can be a costly and time-consuming exercise, it is important for the quality of the survey data that the best interviewers be assigned adequate resources and time be made available so that effective refusal conversion can be achieved.

4. Cost and timeliness

69. The sample size of a survey clearly affects its cost. In general, the overall cost of a survey is a function of fixed overhead costs and the variable costs associated with the selection and processing of each sample unit at each stage of sample selection. Therefore, the larger the sample, the higher the overall cost of survey implementation. A more detailed discussion of the relevant components of the cost of household surveys is provided in chapter XII. Empirical examples of costing for specific surveys are provided in chapters XIII and XIV.

70. The sample size can also affect the time in which the data are made available for analysis. It is important that data and survey estimates be made available in a timely fashion, so that policy decisions can be made on reasonably up-to-date data. The larger the sample, the longer it will take to clean, edit and weight the data for analysis.

F. Survey analysis

1. Development and adjustment of sampling weights

71. Sampling weights are needed to compensate for unequal selection probabilities, for non-response, and for known differences between the sample and the reference population. The weights should be used in the estimation of population characteristics of interest and also in the estimation of the standard errors of the survey estimates generated.

72. The base weight of a sampled unit can be thought of as the number of units in the population that are represented by the sampled unit for purposes of estimation. For instance, if the sampling rate within a particular stratum is 1 in 10, then the base weight of any unit sampled from the stratum is 10, that is to say, the sampled unit represents 10 units in the population, including the unit itself.

73. The development of sampling weights usually starts with the construction of the base weights for the sampled units, to correct for their unequal probabilities of selection. In general, the base weight of a sampled unit is the reciprocal of its probability of selection for inclusion in the sample. In the case of multistage designs, the base weight must reflect the probability of selection at each stage. The base weights for sampled units are then adjusted to compensate for non-response and non-coverage and to make the weighted sample estimates conform to known population totals.

74. When the final adjusted weights of all sampled units are the same, the sample is referred to as self-weighting. In practice, samples are not self-weighting for several reasons. First, sampling units are selected with unequal probabilities of selection. Indeed, even though the PSUs are often selected with probability proportional to size, and households are selected at an appropriate rate within PSUs to yield a self-weighting design, this may be nullified by the selection of one person for interview in each sampled household. Second, the selected sample often has deficiencies including non-response and non-coverage owing to problems with the sampling frame (see sect. C). Third, the need for precise estimates for domains and special subpopulations often requires oversampling these domains (see sect. D).

75 As already mentioned, it is rarely the case that all desired information is obtained from all sampled units. For instance, some households may provide no data at all, whereas other households may provide only partial data, that is to say, data on some but not all questions in the survey. The former type of non-response is called unit or total non-response, while the latter is called item non-response. If there are any systematic differences between the respondents and non-respondents, then naive estimates based solely on the respondents will be biased. To reduce the potential for this bias, adjustments are often made as part of the analysis so as to compensate for non-response. The standard method of compensating for item non-response is imputation, which is not covered in this chapter. See Yansaneh, Wallace and Marker (1998), and references cited therein, for a general discussion of imputation methods and their application to large, complex surveys.

76. For unit non-response, there are three basic procedures for compensation:

- Non-response adjustment of the base weights
- Selection of a larger-than-needed initial sample, to allow for a possible reduction in the sample size due to non-response
- Substitution, which is the process of replacing a non-responding household with another household which was not sampled and which is similar to the non-responding household with respect to the characteristics of interest

77. It is advisable that some form of compensation be used for unit non-response in household surveys, either by adjusting the base weights of responding households or by substitution. The advantage of substitution is that it helps keep the number of participating households under control. However, substitution takes the pressure off the interviewer to obtain data from the original sampled households. Furthermore, attempts to substitute for non-responding households take time, and errors can be made in the process. For example, a substitution may be made using a convenient household rather than the household specifically designated to serve as the substitute for a non-responding household. The procedure of adjusting sample weights for non-response is more commonly used in major surveys throughout the world. Essentially, the adjustment transfers the base weights of all eligible non-responding sampled units to the responding units. Chapter VIII provides a more detailed discussion of non-response and non-coverage in household surveys, and of practical ways of compensating for them (see

also the references cited therein). Chapter XI and the case studies in part two (chaps. XXII, XXIII and XXV) also provide details for specific surveys.

78. Further adjustments can be made to the weights, as appropriate. For instance, if reliable control totals are available, post-stratification adjustments can be employed to make the weighted sampling distributions for certain variables conform to known population distributions. See Lehtonen and Pahkinen (1995) for some practical examples of how to analyse survey data with poststratification.

2. Analysis of household survey data

79. In order for household survey data to be analysed appropriately, several conditions must be satisfied. First, the associated database must contain information reflecting the sample selection process. In particular, the database should include appropriate labels for the sample design strata, primary sampling units, secondary sampling units, etc. Second, sample weights should be provided for each unit in the data file reflecting the probability of selection of each sampling unit and compensating for survey non-response and other deficiencies in the sample. Third, there must be sufficient technical documentation of the sample design for the survey that generated the data. Fourth, the data files must have the appropriate format and structure, as well as the requisite information on the linkages between the sampling units at the various stages of sample selection. Finally, the appropriate computer software must be available, along with the expertise to use it appropriately.

80. A special software program is required to calculate estimates of standard errors of survey estimates that reflect the complexities of the sample design actually used. Such complexities include stratification, clustering and unequal-probability sampling (weighting). Standard statistical software packages generally cannot be used for standard error estimation with complex sample designs, since they almost always assume that the data have been acquired by simple random sampling. In general, the use of standard statistical packages will understate the true standard errors of survey estimates. Several software packages are now available for the purpose of analysis of survey data obtained from complex sample designs. Some of these software packages are extensively reviewed and compared in chapter XXI.

G. Concluding remarks

81. We conclude by emphasizing a few topical issues associated with the design of household surveys in developing and transition countries, namely:

(a) The multi-purpose nature of most household surveys: There is renewed interest, in developing and transition countries, in the establishment of ongoing multi-purpose, multi-subject, multi-round integrated programmes of surveys, as opposed to one-shot, ad hoc surveys. From the outset, the survey designer must recognize the multi-purpose nature of the survey and the competing demands that will be made upon the data generated by it. These competing demands usually impose constraints on the sample that are often very difficult to satisfy. Thus the work of the survey designer should involve extensive discussions with donors, policy makers, data producers at the national statistical office, and data users in the various line ministries of the country. The objective of these preliminary discussions is to attempt to harmonize and rationalize the competing demands on the survey design, before the sample design is finalized;

(b) Determination of an appropriate sample size: One of the major issues to be dealt with at the outset is the determination of an appropriate sample size for a survey. There is increasing demand for precise estimates of characteristics of interest not only at the national and regional levels, but also at the provincial and even lower levels. This invariably leads to demands for large sample sizes. The premium placed on ensuring reliability of survey estimates by reducing sampling error through large sample sizes is far heavier than that placed on the equally significant problem of ensuring data quality by reducing non-sampling errors. It is advisable for the survey designer to perform a cost-benefit analysis of various choices of sample size and allocation scheme. Part of the cost-benefit analysis should involve a discussion of non-sampling errors in surveys and their impact on the overall quality of the survey data. Demands for large sample sizes should be considered only in the light of the associated costs and benefits. As stated in section D, it is important to remember that, in allocating the sample, priority consideration should be given to the domains of interest;

(c) Documentation of the survey design and implementation: For many surveys, documentation of the survey design and implementation process is lacking or insufficient. For a data set to be useful to analysts and other users, it is absolutely essential that every aspect of the design process that generated the data be documented, including the sample selection, data collection, preparation of data files, construction of sampling weights including any adjustments to compensate for sample imperfections and, if possible, specifications for the estimation of standard errors. No appropriate analysis of the data can be conducted without such documentation. Survey documentation is also essential for linkage with other data sources and for various kinds of checks and supplementary analyses;

(d) Evaluation of the survey design: A very important aspect of the survey design process is conducting analyses to evaluate the effectiveness of the design after it is implemented. Resources need to be earmarked for this important exercise as part of the overall budget development process at the planning stage. Evaluation of the current design of a survey can help improve the sample design for future surveys. Such an evaluation can reveal such useful information as whether or not there were any gains from disproportionate allocation; and the extent of the discrepancy, if any, between the current measures of size and those obtained at the time of sample selection. Such information can then be used to develop more efficient designs for future surveys.

Acknowledgements

The author is grateful for the constructive comments of various reviewers and editors, and especially to Dr. Graham Kalton for his numerous suggestions which led to considerable improvements in the initial drafts of this chapter. The opinions expressed herein are those of the author and do not necessarily reflect the policies of the United Nations.

References

Cochran, W.G. (1977). *Sampling Techniques, 3rd ed.* New York: Wiley.

Glewwe, P. and I. Yansaneh (2000). The Development of Future Household Surveys in Viet Nam. Report of Mission to the General Statistics Office, Viet Nam.

Kalton, G. (1983). *Introduction to Survey Sampling*. Quantitive Applications in the Social Sciences Series, Sage University Paper, No. 35. Beverly Hills, California: Sage Publications.

Kish, L. (1965). *Survey Sampling.* New York: Wiley.

_________ (1976). Optima and proxima in linear sample designs. *Journal of the Royal Statistical Society,* Series A, vol. 139, pp. 80-95.

_________ (1988). Multi-purpose sample design. *Survey Methodology*, vol. 14, pp. 19-32.

_________ (1995). Methods for design effects. *Journal of Official Statistics,* vol. 11, pp. 55-77.

Lehtonen, R., and E. J. Pahkinen (1995). *Practical Methods for Design and Analysis of Complex Surveys*. New York: Wiley.

Lohr, Sharon (1999). *Sampling: Design and Analysis*. Pacific Grove, California: Duxbury Press.

Yansaneh, I.S. (2000). Sample Design for the 2000 Turkmenistan Mini-census Survey. Report of Mission to the National Institute for Statistics and Forecasting, Turkmenistan.

__________ (forthcoming). Construction and use of sample weights. *Handbook on Household Surveys.* New York: DESA/UNSD. In preparation.

__________, L. Wallace and D.A. Marker (1998). Imputation methods for large complex datasets: an application to the NEHIS. In *Proceedings of the Survey Research Methods Section, American Statistical Association*. Alexandria, Virginia: American Statistical Association. pp. 314-319.

Annex

Flowchart of the survey process